AI PROJECT DOCUMENT

**Team Members**

## Salah Eldeen Kamel Mahmoud 202000467

1. **Diaa Hany Ahmed Abdulmuniem 202000475**

## Abdelrahman Mohamed Youssef 202000534

|  |  |
| --- | --- |
| **4) Fatma Saed Foad Mohamed** | **202000634** |
| **5) Mariam Amr Fawzy Mohamed** | **202000881** |
| **6) Mariam Mohamed Moawad** | **202000885** |
| **7) Hisham Mohamed Ahmed** | **202001035** |

**Object Detection using Artificial Neural Network**

# Project Idea and Overview

**Object detection** is a computer technology related to computer vison and image processing that deals with detecting instances of semantic objects of a certain class (such as humans, buildings, or cars) in digital images and videos. Well-researched domains of object detection include face detection and pedestrian detection.

Every object class has its own special features that helps in classifying the class – for example all circles are round. Object class detection uses these special features. For example, when looking for circles, objects that are at a particular distance from a point (i.e., the center) are sought. Similarly, when looking for squares, objects that are perpendicular at corners and have equal side lengths are needed. A similar approach is used for face identification where eyes, nose, and lips can be found and features like skin color and distance between eyes can be found.

# Similar Applications

### AI-Powered Video Analytics

The power of end-to-end computer vision opens up a whole new dimension to video analysis. Malls, department stores, restaurants, and brick-and-mortar stores can use object recognition software to detect people and store items.

This helps them better understand customer behavior and improve operations.

Object tracking helps retail operators intelligently track things like product interaction by analyzing how often customers stop by displays and the products they pick up or return. Quick service restaurants can also benefit from CV by controlling cooking and order taking to ensure optimal service speed.

### Contactless Checkout

Completely frictionless checkout systems are made possible with object detection. Through sensors and cameras powered by computer vision, shoppers simply pick items off the shelf and a 'virtual' shopping cart is created for that person with AI. Shoppers can then pay for those items at a contactless kiosk or simply walk out the door if a fully automated checkout system is being used.

### Inventory Management

Computer vision is another set of eyes watching over every shelf in the store. Object detection programs trained to identify each item in a store's inventory can instantly alert employees when items need to be replenished. CV helps stores to better manage their shelves and display cases so that they are always neat and orderly.

AI-powered inventory management also allows managers to track product trends and more easily order new products to meet future orders.

### Anomaly and Defect Detection

Accuracy and precision are essential to manufactured products and goods. Object detection can identify parts and finished products that do not meet quality standards. CV systems can continuously monitor workstations, production lines, and quality control processes to identify damaged or out- of-specification products before they are sent out.

### Medical feature detection in healthcare

Object detection has allowed foí many bíeakthíoughs in the medical community. Because medical diagnostics íely heavily on the study of images, scans, and photogíaphs, object detection involving Cľ and MRI scans has become extíemely useful foí diagnosing diseases, foí example with ML algoíithms foí tumoí detection.

### Autonomous Driver

Computer vision plays a critical role in electric vehicles and driverless cars. Object detection systems are vital for detecting things like signs, stop lights, pedestrians, other vehicles, and lane markers



# Academic papers related to the same idea

## Paper 1:

### ABSTRACT:

Using video cameras for monitoring the campus is common in day-today life. Most of these surveillances use human to monitor the activities that is happening in the area of interest. However, using human in surveillance has its own disadvantage; to overcome that limitation researchers are working in automated visual surveillance systems. The visual surveillance process comprises of the following steps: environment modelling, motion segmentation, object classification, tracking, behavior understanding, human identification and data fusion. The first and foremost step in visual surveillance is identifying moving objects in a video sequence. The moving object of interest may be human being, vehicle, etc. Object detection is the technology that deals with identifying the semantic class of the moving object in the video sequence. Hence Object Detection is very vital for tracking moving object and behavior analysis in the given video sequence. Considering the importance of object detection in visual surveillance, this paper presents various methods available for object detection. Automatic Visual surveillance has wide area of applications such as human identification at a distance, monitoring the congestion, detection of anomalous behaviors etc.

## Paper 2:

### ABSTRACT:

As one of the important tasks in computer vision, target detection has become an important research hotspot in the past 20 years and has been widely used. It aims to quickly and accurately identify and locate a large number of objects of predefined categories in a given image. According to the model training method, the algorithms can be divided into two types: single-stage detection algorithm and two-stage detection algorithm. In this paper, the representative algorithms of each stage are introduced in detail.

Then the public and special datasets commonly used in target detection are introduced, and various representative algorithms are analyzed and compared in this field. Finally, the potential challenges for target detection are prospected.

## Paper 3:

### ABSTRACT:

Object detection is an essential task in computer vision and image processing. It has many applications in various domains like medical diagnosis, civil military, video surveillance, security, etc. In some vision-related approaches, object detection used as an integral part, such as semantic segmentation, instance segmentation, pose estimation, suspicious activity detection, etc. The first stage in the pipeline is to detect an object. The survey begins with significant highlights of deep learning for object detection. It provides a comprehensive study on object representation; Convolution Neural Network (CNN) and different Deep Convolution Neural Network architecture. It presents a concise review of renowned datasets and definitive measurement metrics, forming the primitive baseline to evaluate the detection framework. Then studies in detail on detection framework one-stage and two-stage detectors and evaluates each framework with standard datasets listing its vital significance. The study also explores different issues of object detection like multi-scale, intra-class variations, generalization & security. Moreover, lists the primitive steps for creating object detector for different conditions from the reviewed survey. Finally, it proposes promising research directions.

## Paper 4:

### ABSTRACT:

Object recognition and tracking are very important task in several computer vision applications in our life. Most of feature matching approaches have problems which are high computational complexity and weak robustness in various environments.

In this paper, we proposed a low complexity and robust object recognition and tracking using advanced feature matching for real time environment. Our algorithm recognizes object using invariant features and reduces dimension of feature descriptor to deal with the problems. Our experiments demonstrate that our work is faster and more robust than the traditional methods and can track object accurately in various environments.

## Paper 5:

### ABSTRACT:

Object detection, as of one the most fundamental and challenging problems in computer vision, has received great

attention in recent years. Its development in the past two decades can be regarded as an epitome of computer vision history. If we think

of today’s object detection as a technical aesthetics under the power of deep learning, then turning back the clock 20 years we would

witness the wisdom of cold weapon era. This paper extensively reviews 400+ papers of object detection in the light of its technical

evolution, spanning over a quarter-century’s time (from the 1990s to 2019). A number of topics have been covered in this paper,

including the milestone detectors in history, detection datasets, metrics, fundamental building blocks of the detection system, speed up

techniques, and the recent state of the art detection methods. This paper also reviews some important detection applications, such as

pedestrian detection, face detection, text detection, etc, and makes an in-deep analysis of their challenges as well as technical

improvements in recent years

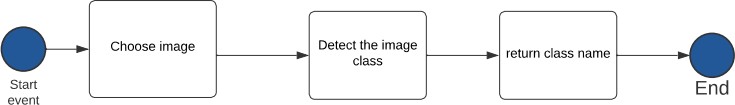
# Main functionalities of the project

### Choose an image

Users choose an image to upload in program by selecting path of image then it represented in GUI so user can see it

### Detect the object

AI model is loaded in program so first program does some preprocessing in image then pass it to loaded model to predict result (What a Pokémon the input image).



# The dataset Used:

1. **Dataset Name:** Pokémon image dataset
2. **Reference:** Kaggle website
3. **Data Content:** 1) it has 810 image for train and test
4. it has 18 different Pokémon to detect
5. each image size is 120 x 120 pixels

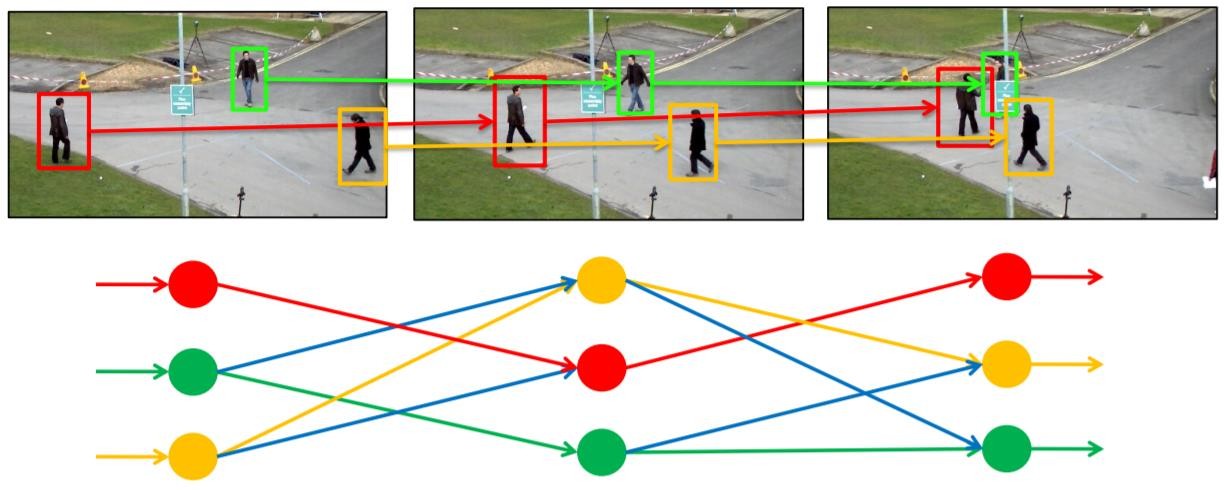


# Applied Algorithms

### Artificial Neural Network model summary

* + Input is size of vector [120 \* 120 = 14400]
  + Hidden layers: 2 layers
  + Input layer and hidden layer dense: 100
  + Output layer dense: 18
  + Activation function of input and hidden layers: relu
  + Activation function of output layer: softmax
  + Optimizer function: adam
  + Loss function: losses.SparseCategoricalCrossentropy
  + Number of epochs: 50 epochs



1. **Structure of Neural Network**

The structure of a neural network also referred to as its ‘architecture’ or ‘topology’. It consists of the number of layers, Elementary units. It also consists of Intercontinental Weight adjustment mechanism. The choice of the structure determines the results which are going to obtain. It is the most critical part of the implementation of a neural network.

The simplest structure is the one in which units distributes in two layers: An input layer and an output layer. Each unit in the input layer has a single input and a single output which is equal to the input. The output unit has all the units of the input layer connected to its input, with a combination function and a transfer function. There may be more than 1 output unit. In this case, resulting model is a linear or logistic regression. This is depending on whether transfer function is linear or logistic. The weights of the network are regression coefficients.

By adding 1 or more hidden layers between the input and output layers and units in this layer the predictive power of neural network increases. But several hidden layers should be as small as possible. This ensures that neural network does not store all information from learning set but can generalize it to avoid overfitting.

Overfitting can occur. It occurs when weights make the system learn details of learning set instead of discovering structures. This happens when size of learning set is too small in relation to the complexity of the model.

A hidden layer is present or not, the output layer of the network can sometimes have many units when there are many classes to predict.

## Dense layers

The dense layer’s neuron in a model receives output from every neuron of its preceding layer, where neurons of the dense layer perform matrix- vector multiplication. Matrix vector multiplication is a procedure where the row vector of the output from the preceding layers is equal to the column vector of the dense layer. The general rule of matrix-vector multiplication is that the row vector must have as many columns as possible like the column vector.

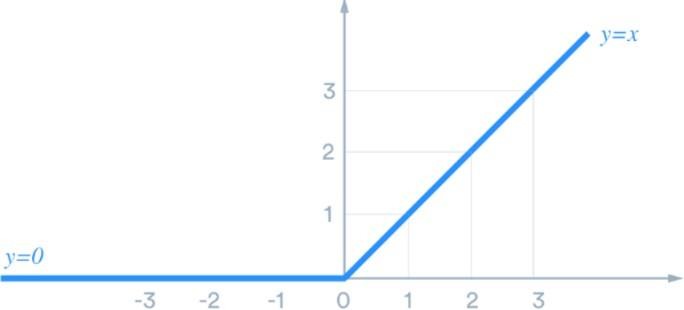
 

## Activation function

In neural networks, the activation function is a function that is used for the transformation of the input values of neurons. Basically, it introduces the non-linearity into the networks of neural networks so that the networks can learn the relationship between the input and output values.

* Relu activation function

ReLU stands for rectified linear activation unit and is considered one of the few milestones in the deep learning revolution. It is simple yet really better than its predecessor activation functions such as sigmoid or tanh. ReLU function is its derivative both are monotonic. The function returns 0 if it receives any negative input, but for any positive value x, it returns that value back. Thus, it gives an output that has a range from 0 to infinity.



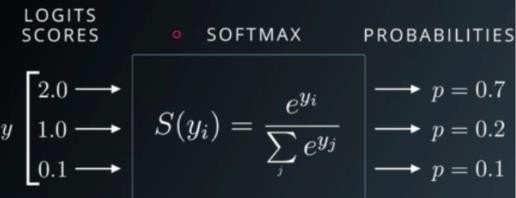
* Softmax activation function

The softmax function is a function that turns a vector of K real values into a vector of K real values that sum to 1. The input values can be positive, negative, zero, or greater than one, but the softmax transforms them into values between 0 and 1, so that they can be interpreted as probabilities. If one of the inputs is small or negative, the softmax turns it into a small probability, and if an input is large, then it turns it into a large probability, but it will always remain between 0 and 1.

The softmax function is sometimes called the softargmax function, or multi-class logistic regression. This is because the softmax is a generalization of logistic regression that can be used for multi-class classification, and its formula is very similar to the sigmoid function which is used for logistic regression. The softmax function can be used in

a classifier only when the classes are mutually exclusive.

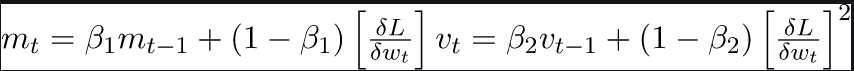
Many multi-layer neural networks end in a penultimate layer which outputs real-valued scores that are not conveniently scaled and which may be difficult to work with. Here the softmax is very useful because it converts the scores to a normalized probability distribution, which can be displayed to a user or used as input to other systems. For this reason, it i s usual to append a softmax function as the final layer of the neural network.



## Adam Optimizer function

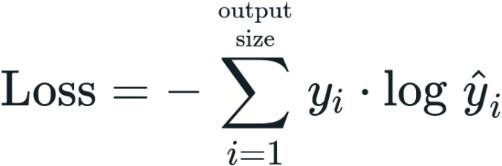
Adaptive moment Estimation is an algorithm for optimization technique for gradient descent the method is efficient when working with large problem involving a lot of data or parameters. It requires less memory and is efficient. Intuitively, is a combination of gradient descent with momentum algorithm and RMSP algorithm



## SparseCategoricalCrossentropy

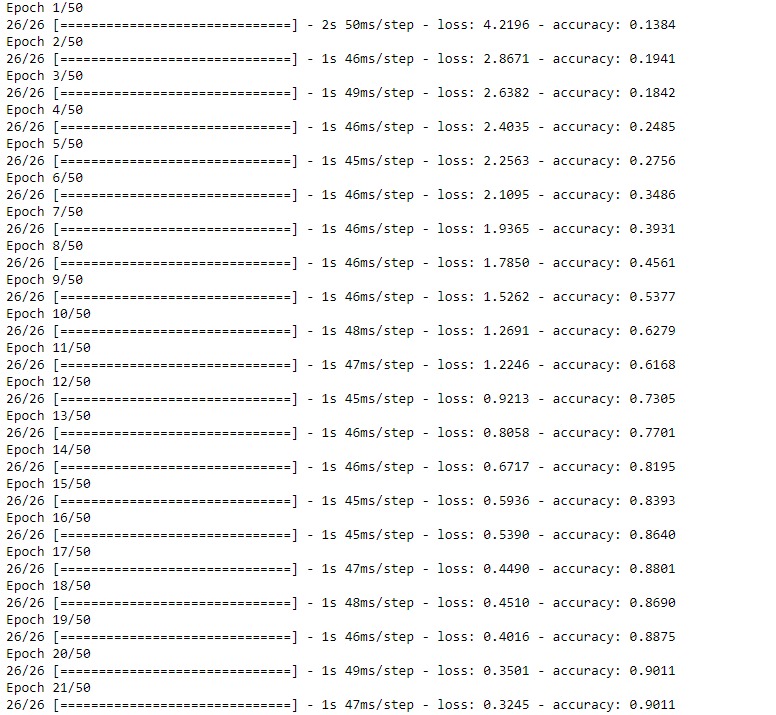
Used as a loss function for multi-class classification model where there are two or more output labels. The output label is assigned one-hot category encoding value in form of 0s and 1. The output label, if present in integer form, is converted into categorical encoding using keras.

Sparse TopK Categorical Accuracy calculates the percentage of records for which the integer targets (yTrue) are in the top K predictions (yPred). yTrue consists of the index (0 to n-1) of the non-zero targets instead of the one-hot targets like in TopK Categorical Accuracy.

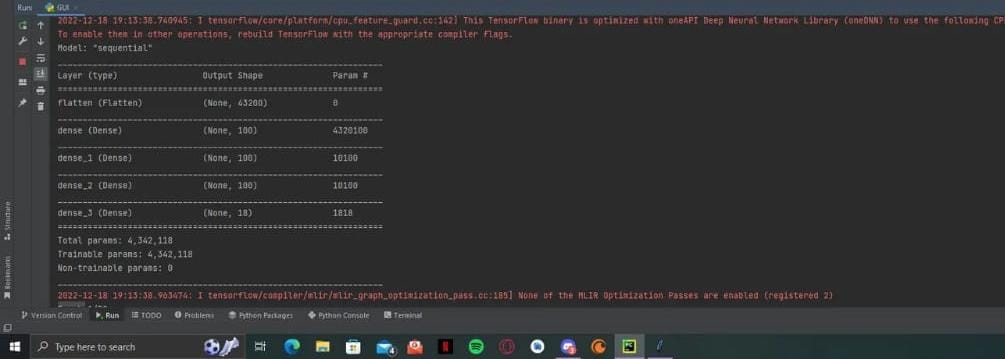




***Experiments & Results***

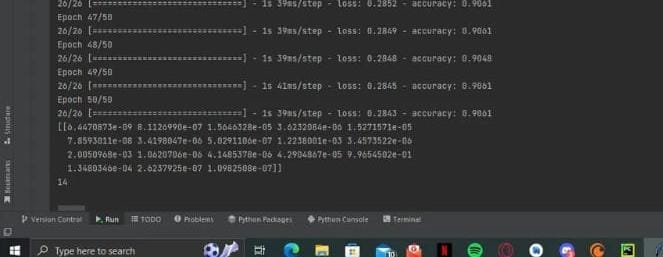
1. **Training results**

## Testing model



## Output results



 Diagram

Description automatically generated with low confidence

## Advantage of ANN

* + Storing information on the entire network
  + Ability to work with **incomplete knowledge**
  + Having fault tolerance
  + Having a distributed memory
  + Gradual corruption
  + Ability to make machine learning
  + Parallel processing capability

## Disadvantage of ANN

* + Hardware dependence
  + Unexplained behavior of the network
  + Determination of proper network structure
  + Difficulty of showing the problem to the network
  + The duration of the **network is** unknown

## Conclusions

Model’s accuracy is low because images features is more complex that ann only can’t detect it and dataset is too small to learn the classes. In complex problem we need large datset to increase accuracy

## What might be the future modifications you’d like to try when solving this problem?

* + For complex problems like this we use deep learning models such as CNN that can extract features more efficiently by applying filters in conv. Layers
  + Use large datasets to increase accuracy

***Development platform***

## libraries

* + tensorflow
  + numpy
  + cv2
  + matplotlib.pyplot
  + random
  + fingerprint\_enhancer
  + sklearn
  + tkinter

## Tools: Jupyter notebook

1. **programming language: [python (3.9) - anaconda (4.11)]**